

基于 VoiceXML 的火车时刻语音查询系统^①

The Train Schedule Speech Inquiring System Based on VoiceXML

陶利民 (杭州师范学院 杭州 310036)

摘要:首先主要介绍了 VoiceXML 的体系结构及工作原理,然后介绍了一个具体的应用实例—火车时刻语音查询系统。

关键词:VoiceXML 语音技术 语音查询

1 引言

VoiceXML (Voice Extensible Markup Language, 语音扩展标记语言) 是一种基于 XML 的因特网标记语言,是一个新的 XMLschema, 用来制定通过语音对话访问 Web 的内容及其交互语音应答的传递标准,从面向应用的角度来说,也就是开发语音用户界面。它是“语音 Web”使用的语言,它使得用户可以使用电话来访问因特网托管的内容, VoiceXML 则通过语音界面来访问网站内容。

2 VoiceXML 的结构模型

VoiceXML 的结构模型如图 1 所示,由四个功能模块组成。

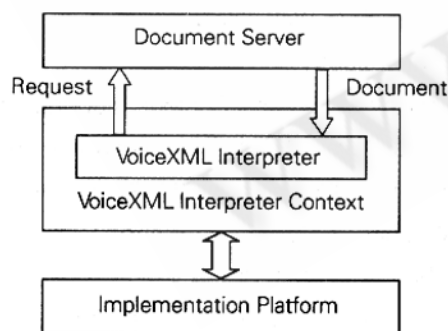


图 1 VoiceXML 结构模型

一个文档服务器(如一个 Web 服务器),处理一个

来自客户应用的请求,即 VoiceXML 解释器,经过 VoiceXML 解释器上下文环境处理。作为响应,服务器产生出 VoiceXML 文档,返回 VoiceXML 解释器的处理。VoiceXML 解释器上下文环境负责监视与 VoiceXML 解释器平行处理的用户输入。实现平台受 VoiceXML 解释器上下文环境和 VoiceXML 解释器控制。

2.1 VoiceXML Interpreter (VoiceXML 解释器)

VoiceXML 解释器的功能类似于计算机上使用的浏览器,因此也称为语音浏览器。语音浏览器是驻留在网络上的一种应用程序,它位于语音、电话和因特网组件之上,负责将 VoiceXML 应用程序传递给用户。语音浏览器由负责解析 VoiceXML 文档的 VoiceXML 翻译器以及连接语音和电话组件的界面组成。语音浏览器通过 VoiceXML 解释器上下文从 Document Server (文档服务器)上读取 VoiceXML 文件,并解释执行。调用 VoiceXML 解释器上下文和实现平台中的 ASR 语音识别引擎、TTS 文本转语音引擎、电话平台共同完成 VoiceXML 文件的处理流程,如播放提示音、等待语音应答、识别语音命令和键盘 DTMF 输入、响应事件等。VoiceXML 解释器也支持客户端脚本和对象处理。

2.2 VoiceXML Interpreter Context (VoiceXML 解释器上下文)

VoiceXML 解释器上下文与 VoiceXML 解释器的关系等同于计算机中操作系统和 Web 浏览器的关系。也就是说, VoiceXML 解释器上下文是 VoiceXML 解释器的运行支持环境。VoiceXML 解释器通过 VoiceXML 解

^① 基金项目:浙江省教育厅科研项目(20050718)

释器上下文与实现平台、文档服务器联系。它包含 VoiceXML 解释器所需要的所有支撑功能。

2.3 Document Server(文档服务器)

文档服务器起到存放 VoiceXML 文件和相关资源以及响应 VoiceXML 解释器的文件请求功能。实际系统中一般由 Web 服务器担任, VoiceXML 解释器通过 HTTP 协议向 Document Server 请示 VoiceXML 文件。因此, Document Server 可以放在远离实现平台的其他地方, 只要网络可达即可。但一般从响应速度和安全角度考虑, Document Server 和实现平台放在同一个本地网中比较好。

2.4 Implementation Platform(实现平台)

实现平台是指具体的硬件平台和网络设备及软件, 包括电话平台、网络设备、语音识别 ASR 系统和语音合成 TTS 系统。负责处理电话网连接、电话呼叫处理、DTMF 输入、语音识别和语音合成等功能。

3 VoiceXML 的工作原理

语音技术不但让那些由于环境或生理限制无法使用图形化浏览器的人得以访问 Web, 也为所有的用户提供了更为便捷的 Web 访问功能。新的语音技术可以创建出对话驱动的应用程序, 例如语音识别技术(ASR), 语音合成技术(TTS)以及记录和回放数字化语音在 PC 和服务端(分发到客户端设备)上。VoiceXML 提供了一种与技术无关的语言, 可以用在语音应用程序。这些应用程序将后端处理语音的服务和处理机制与前端基于 Voice XML 的表示层分离开。例如, 一个精心设计的 Web 站点可以轻而易举地支持语音驱动的浏览器(比如您很可能在移动电话上会使用的那种), 同时它有能支持其它的浏览器(比如一个 WAP 浏览器或 HTML 浏览器)。当接受到来自浏览器的最初请求时, 服务器将监测出浏览器的类型。如果浏览器被确认为一个语音浏览器, 服务器将返回对应的 Voice XML 页面。VoiceXML 的工作原理如图 2 所示。

从图中我们可以看出, 同传统的 Internet 网站相比, 在不需要太大的改动前提下, 增加一个 VoiceXML 的服务器来处理来自 PSTN(公用电话交换网)的应用即可。在 VoiceXML 的服务器中, 包含了 VoiceXML 的解释器(VoiceXML 解释程序语境), VoiceXML 自带的浏览器, 自动语音识别(ASR)及文本到语音(TTS)的转换

设备。VoiceXML 解释器是一个计算机程序。它解释一个 VoiceXML 文件, 引导和控制用户与实现平台之间的交互作用。VoiceXML 解释程序语境也是一个计算机程序, 用一个 VoiceXML 程序解释一个 VoiceXML 文件, 并且可以与实现平台相互作用而与 VoiceXML 解释程序无关。电话平台增加对网上资源进行解释的一个服务模块、对人说话进行识别及对文本内容进行语音做成的模块, 这个模块由解释语境和解释程序核心部分构成, 实现电话能够畅游互联网, “阅读”网上的信息, “点击”网上的链接。VoiceXML 为语音和电话应用的开发者、服务提供商和设备制造商提供了一个智能化的 API, 使人们能够通过语音和电话访问网站上的信息和服务。

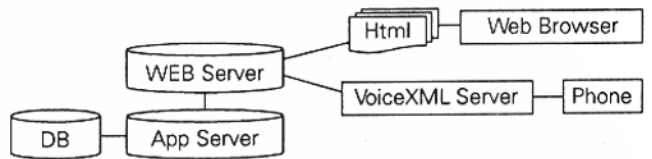


图 2 VoiceXML 工作原理

一次服务的过程为:

- (1) 由解释器等待电话用户的接入, 如果有电话连接到 Vxml 文档服务中, 平台就会通知解释器, 并将相关信息通知解释器;
- (2) 解释器启动解释器核心, 解释器核心向文档服务器获取 Vxml 文档;
- (3) 解释器核心解释文档语义, 根据语义和用户发生交互, 其中包括通过平台向用户播放语音文件或者播放文本合成语音, 获得用户的按键输入或者语音输入;
- (4) 文档结束, 解释器核心结束工作。

4 VoiceXML 的特点和应用范围

4.1 VoiceXML 的特点

VoiceXML 作为一种标记语言, 具有以下几个特点:

- (1) VoiceXML 作为一种通过每一文件里指定的多重的交互作用, 最小化客户机/服务器之间的交互工作;
- (2) 实现应用开发者与低层的软件和系统平台上的软、硬件细节无关;

(3) 将用户交互作用的代码(在 VoiceXML 中)从服务逻辑(CGI 脚本)中分离出来;

(4) 能跨越不同的执行平台,促进服务的可移植性。对于内容服务商、工具提供商和平台提供商来说, VoiceXML 是一个公共语言(适用性较强);

(5) 非常容易用于简单的(人机)交互,要求所提供的语音界面能支持复杂的对话;

(6) 以应用、会话、文档为单位建立应用结构;

(7) 以对话为交互单位,逐个完成对话,确定流程导向;

(8) 与范围相关的语法激活/禁止机制;

(9) 以语音网页为单位,构建复杂的应用层次。

4.2 XML 的应用范围

VoiceXML 语言描述了语音应答系统的人机交互过程,其范围包括:合成语音的输出(TTS)、音频文件的输出、语音输入的认识、DTMF 输入的认识、语音输入的录音、对话流的控制、电话功能(如呼叫转移和挂断等)。VoiceXML 主要用于电话语音系统的开发,具体来说,我们可以把 VoiceXML 应用在以下的实际领域:

(1) 信息的获取。将信息分类提供给用户,通过电话以语音方式获取,如公用信息:新闻、体育赛况、股票信息、天气预报、航班信息等。

(2) 电子商务。电子商务的应用可以提供对电话的支持,也称为语音电子商务。可提供商务包括:订机票、订车票和购物等。

(3) 语音门户。它是一个访问其他语音信息的统一入口,还可提供一些公共服务,例如,用户只需用普通电话机或移动电话拨打语音门户网的特定号码,就可在语音提示引导之下用自然语言和耳朵便捷地“浏览”天气、新闻、体育等多方面的信息。在这个过程中,不必键入菜单选择,仅需简单的语音应答即可完成操作。

(4) 电信领域的服务。如用电话发 E-mail 业务是指用户通过电话接入系统,用户打电话告知系统所要发信的目的 IP 地址,由 ASR 自动识别系统内定义好的 IP 地址,通过 IVR 节点录音的语音文件以附件的方式生成 Email,通过 Internet 发关到对方的 Email 信箱中。还有,UNIFIEDMESSAGE, CALLCENTER(呼叫中心)等服务。当然,我们可利用 VoiceXML 来实现火车时刻的

语音自动查询,包括发站、到站时刻的查询、班次的查询、票价的查询等功能。

5 火车时刻语音查询系统

电话订票方式与传统的窗口排队买票方式并行将为火车票的订售带来极大的便利。也许大家使用过传统的电话订票系统服务,可以实现业务咨询、火车票预订、投诉建议等功能,可以说系统的业务流程比较清晰,也有良好的导航性,但缺点就是互动性较差,且只能有一种输入方式。火车时刻自动语音查询系统的语音平台采用 3W 认可的最新语音流程语言——VOICEXML,不但实现了语音流程的个性化,还可以通过简单地编写 VOICEXML 流程,轻而易举地生成各种各样的语音应用系统,实现平台与业务的分离。

5.1 TSSIS 系统的体系结构

TSSIS 系统的体系结构如图 3 所示。

TSSIS 系统为了保证系统的可靠性,在提供电话服务的同时,也提供传统的 Web 网络服务。图 3 中系统的上半部分使用了传统的 Web 应用体系结构。GUI 浏览器通过对 HTML 文档进行解析和处理向用户提供可视化信息(文本、图像、声音、超链接等);当用户做出了某种选择(如点击了某个超链接)时,GUI、浏览器会向 Web 服务器发送一个 HTTP 请求(即检索另一个 Web 页面),Web 服务器调用新的页面并将该页面的 HTML 内容发送给浏览器。此外,Web 服务器还需要与后端系统(数据库、Servlets 等)交互来获得必要的信息。

应用 VoiceXML 的语音系统则对上述 Web 访问模式进行了扩展。图 3 中系统的下半部分添加了一个电话(固定电话、移动电话、PDA)、一个 PBX 和一个 VoiceXML 服务器。用户拿起电话通过 PSTN 提交查询信息。交换机(PBX)实现电话呼入、呼出功能;还提供自动排队机系统,在来电量小时用于实现对接话的均衡分配。VoiceXML 服务器用于管理语音浏览会话(每个呼叫对应一个会话),而每个语音浏览会话中又包括一个语音浏览器实例、一个语音识别引擎实例和一个 TTS 引擎实例。

VoiceXML 提供了一种表示 Web 信息的新方式。在这种方式中,语音浏览器不是通过 HTML 那样提供图像、文本等可视化信息,而是通过 VoiceXML 向用户提供音频信息。当呼叫者发出语音指令时(与点击某

个超链接类似), VoiceXML 服务器会向 Web 服务器发送一个 HTTP 请求, 然后 Web 服务器通过与后端系统进行交互便可以对该 HTTP 请求进行响应, 不过这次是以 VoiceXML 的形式进行响应。

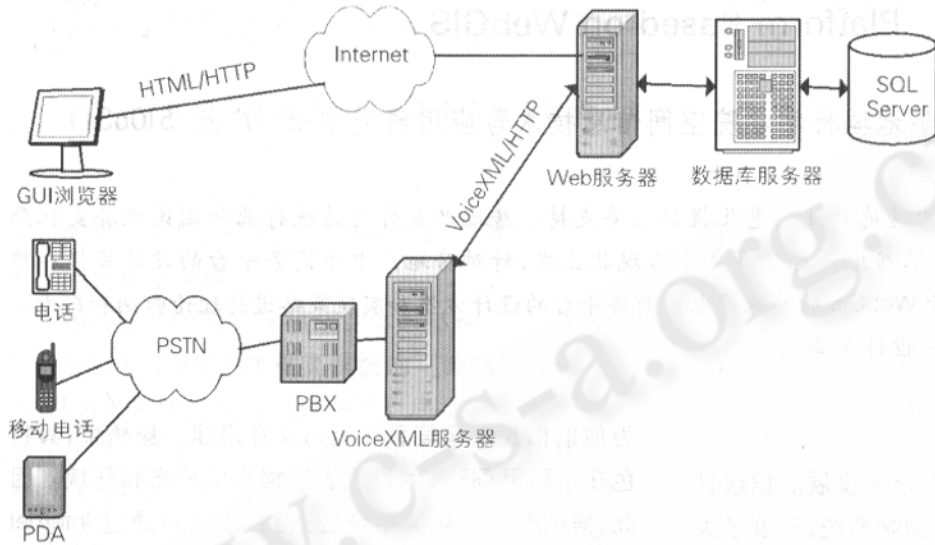


图 3 TSSIS 网络体系结构

VoiceXML 服务器启动时, 会通过局域网或 Internet 向 Web 服务器发送一个 HTTP 请求以获得一个初始 VoiceXML 文档, 该文档可以包含静态的信息, 也可以像动态 HTML 那样, 使用同样的服务器端逻辑 (CGI 脚本、Java Bean、ASP、JSP、Java Servlet 等) 对企业数据库进行访问并动态地生成 VoiceXML 文档。VoiceXML 服务器对该 VoiceXML 文档进行解析和处理, 根据用户的输入, 它会确定是向 Web 服务器请求新的 VoiceXML 文档, 还是让 Web 服务器更新后端数据库中的数据。这种访问方式优点在于, 我们并不需要改变后端企业数据的访问机制, VoiceXML 应用可以访问 HTML 应用所能访问的同样信息。

数据库服务器存储火车时刻数据、用户资料、服务记录等各类信息, DBMS 采用 SQL Server2000。鉴于对系统的安全性、可靠性的考虑, 系统使用 Web Service 接口访问后台的业务处理系统。

5.2 系统工作过程

通过电话来查询火车时刻, 其工作过程如下:

- (1) 用户拨打 96111 (假设这是某火车站客户服务程序的电话号码);
- (2) 通过公用电话交换网 (PSTN) 连接到负责托

管该服务程序的语音服务提供商 (VSP) 的 VoiceXML 网关, 这里说的连接实际上就是语音浏览器的入口, 相当于 HTML 浏览器的地址栏;

- (3) 语音浏览器在获知呼叫号码和主叫号码的信息后就去找搜索一个数据库, 同时将呼叫号码转换为一个统一资源地址 (URL) 这个 URL 就是用 VoiceXML 语言开发的移动客户服务程序 ClientServe.vxml 的位置;

(4) 语音浏览器激活 ClientServe.vxml 主页, 如果预先进行了设定, 那么也可以激活相关的文件, 它们被从 Web 服务器上下载并高速缓存在 VoiceXML 网关上;

(5) 语音浏览器对 VoiceXML 文档进行翻译, 根据流程执行 ClientServe.vxml, 并根据应用程序的规定与用户

之间进行交互操作, 应用程序输入部分可能要借助于 ASR, 输出要调用 TTS;

- (6) 根据 ClientServe.vxml 需要, 从 Web 服务器上下载其它的 VoiceXML 文件并继续流程。

5.3 系统应用

(1) 查询与咨询服务。用户可在语音菜单提示下查询火车时刻、票价、车票分布情况等信息, 如需要, 可随时请求人工服务。

(2) 业务受理服务。业务受理分为自动业务受理和人工受理, 业务受理包括系统维护、发送传真、收费确认等功能, 用户可根据需要自动转换人工还是自动处理。

(3) 投诉建议。投诉建议提供人工和自动两种受理方式, 投诉和建议分开处理, 投诉主要受理前台服务、服务质量、新业务使用、终端维修等有关的投诉。建议主要受理对铁路公司新老业务、网络质量、前台服务等各类需求建议。

(4) 信息发布、服务分类统计和分析。火车站可利用呼叫中心及时通知用户最新的信息, 和用户取得

(下转第 17 页)

(上接第 13 页)

联系,并且可以对铁路公司的各项服务进行分析和统计,及时得到用户反馈,从而能进一步提高服务质量。

(5) 语音门户网站。VoiceXML 先进的技术可以将传统的铁路公司的网站改造为语音门户网站,用户既可以通过传统的 Internet 方式浏览铁路公司网站,也可以通过电话拨入的方式获得网站上的信息,使用网站上的功能,利用 JSP 的技术,实现语音服务数据和网上浏览数据同步更新。

6 结束语

语音应用系统利用现有发达的电话网,使用户能够以最简单、最方便、最满意的方式获取信息。以“说

话”的方式来操纵计算机网络系统,将彻底改变人们几十年形成的按键操作的传统概念和习惯,使人重归自然,轻松自由。

参考文献

- 1 王建华、王卫峰等译, VoiceXML 语音应用程序开发 [M], 北京:机械工业出版社,2002.
- 2 吕俊辉、邓时颖、李怀诚, VoiceXML 规范[J], 信息技术与标准化,2002 年第 9 期.
- 3 杨广润、汪恒杰、崔子筠, 基于 VoiceXML 的呼叫中心系统的实现[J], 计算机应用研究,2003. 6.